DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

MEMORANDUM

Food and Drug Administration
10903 New Hampshire Ave
Silver Spring, MD 20993-0002

Date: October 8, 2009

From: Mary Elizabeth Ritchey, RN, MSPH, PhD
Cara Krulewitch, CNM, PhD, FACNM
Hui Lee Wong, PhD, MS
Office of Surveillance and Biometrics (OSB)/Division of Epidemiology (DEPI)

Subject: Epidemiologic Review: Lung and Colon CADe

To: Janine Morris, Acting Director, DRARD/ODE

Through: Danica Marinac-Dabic, MD, PhD, Director, OSB/DEPI ______

Purpose:
The purpose of this memorandum is to present an epidemiologic review of the scientific literature on the use of computer assisted detection/diagnosis (CADe) for detection of pulmonary nodules and colonic polyps as part of screening programs.

I. Introduction

More people die from lung cancer in the U.S. than any other type of cancer. This is true for both men and women. In 2005 (the most recent year for which statistics are currently available), lung cancer accounted for more deaths than breast cancer, prostate cancer, and colon cancer combined. In 2005, 107, 416 men and 89,271 women were diagnosed with lung cancer and 90, 139 men and 69,078 women died from lung cancer (CDC, 2009).[1] There is some evidence that early detection of lung cancer may result in a more favorable prognosis.[2] This has led some to propose lung cancer screening in high-risk patients using chest computed tomography (CT) scans. This remains a topic of debate in the literature.[3]

Colon cancer is one of the most important forms of cancer, as it is the third most common cancer and the second greatest cause of mortality from cancer in the United States. According to the Surveillance, Epidemiology and End Results program of the National Cancer Institute, a total of 146,970 Americans (75,590 men and 71,380 women) will be diagnosed with and 49,920 Americans will die of cancer of the colon and rectum in 2009.[4]

The vast majority of colon cancers arise from colonic polyps. The timely detection and removal of colonic polyps can prevent the development of colon cancer. The value of secondary prevention for colon cancer is heightened by the particularly poor prognosis of metastatic colon cancer. Patients with metastatic spread of colon cancer to other organs, such as the liver, have a 5-year survival rate of less than 10%.

Colonic polyps are generally asymptomatic and cannot be detected by routine physical examination. The current gold standard for the detection and removal of polyps is the use of optical colonoscopy (OC). This procedure requires the insertion and advancement of a flexible endoscope through the length of the colon. Disadvantages of this procedure include the need for sedation, patient discomfort, a small, but real risk of colonic perforation with adverse sequellae including death, and a risk of infectious disease transmission. Flexible endoscopes cannot tolerate steam sterilization and their long, narrow channels are challenging to adequately clean. The absence of adequate cleaning can compromise the disinfection process. Computed tomography (CT) colonography, also known as virtual colonoscopy, has been developed as a potential non-invasive alternative to optical colonoscopy.

CADe systems for both lung cancer screening and CT colonography have been introduced into clinical practice as a second reader (after the physician has performed an initial interpretation of the patient chest data) to aid/assist the physician's identification/detection of solitary/solid pulmonary nodules or potential polyps. The purpose of this memorandum is to provide an assessment of the risks and benefits of the addition of CADe systems to these procedures when used as screening tools.

II. Methods

A search of the MEDLINE database was performed in PubMed using the following terms that define CADe: computer AND (assisted, -assisted, aided, -aided, based, or -based) AND (detection, diagnosis or diagnoses). This search yielded 52,483 abstracts.

A search to define lung involvement was conducted using the following Mesh terms: 'lung', 'lung neoplasms', 'lung diseases', 'pleura', 'solitary fibrous tumor, pleural', 'pleural cavity', 'pleural neoplasms', 'pleural diseases', 'thorax', 'thoracic cavity', 'multiple pulmonary nodules'. These terms were added to the non-Mesh search: (nodule and (lung or pulmonary)). The combined search yielded 748,419 abstracts.

A search to define colon involvement was conducted using the following Mesh terms: 'colon', 'colon, descending', 'colon, sigmoid', 'colonic neoplasms', 'sigmoid neoplasms', 'sigmoid Diseases', 'colorectal neoplasms, hereditary nonpolyposis', 'colonic polyps', 'colonoscopy', 'colonography, computed tomographic', or 'sigmoidoscopy'. This search yielded 113,410 abstracts.

We combined the CADe and lung searches and restricted the search to studies published after 2007 in English. We required the term "CADe" to appear in the title or abstract. This combined search yielded 159 relevant abstracts.

We combined the CADe and colon searches and restricted the search to studies published after 2007 in English. We required the term "CADe" to appear in the title or abstract. This combined search yielded 80 relevant abstracts.

The criteria for inclusion were all prospective or retrospective studies that compared the addition of CADe to routine radiological screening techniques: (1) as part of routine screening (with all cases included), and (2) studies that selected cancer cases and controls and evaluated reader behavior (reader studies). Studies in children were not considered.

After review, 4 lung abstracts and 2 colon abstracts were considered relevant and full texts were ordered. The 4 lung and 2 colon abstracts represented 6 unique clinical evaluations that were relevant to this literature review.

III. Lung: Retrospective Reader Performance and Clinical Studies

Kasai et al[5] reported the change in detection for 18 radiologists assessing posteroanterior and lateral chest images of 21 patients with vertebral fractures, 31 patients with lung nodules, and 10 control patients. The goal of this study was to evaluate the usefulness of CADe in detection of vertebral fractures and lung nodules. The fracture cases were considered severe by consensus of two radiologists prior to the study and the lung nodules ranged 8.3-29.8 mm. The authors calculated the sensitivity of screening with and without CADe, as well as the overall accuracy for both vertebral fracture and lung nodule detection, measured using both area under the ROC curve and jackknife free-response ROC. With the addition of CADe, the average area under the ROC curve improved for both detection of vertebral fractures (0.906 to 0.951, $p = 0.002$) and lung nodules (0.804 to 0.816, $p = 0.297$), though the difference for lung nodules was not statistically significant. The values obtained with jackknife free-response ROC improved from 0.585 to 0.680 ($p < 0.001$) for vertebral fractures and from 0.622 to 0.650 ($p = 0.017$) for nodules. Average sensitivity of detection improved from 59.8% to 69.3% for vertebral fractures and from 64.9% to 67.6% for nodules. The authors concluded that the diagnostic accuracy of vertebral fractures and lung nodules increased with CADe. They found that CADe had the most benefit in detection of fractures below the diaphragm or in the upper lung area and in detection of nodules overlapping a vertebral body or the clavicle or close to the heart.

Roos et al[6] conducted a study in which three radiologists assessed chest CT from 20 patients with clinical suspicion of pulmonary nodules. The objective of the study was to assess the temporal variation in performance of radiological evaluations using incremental CADe assistance. CADe was added after an initial CT assessment and CADe detections were added individually in order of their likelihood of being a nodule. The average sensitivity for the initial search was 53% with 1.15 false positives (FP) per patient. The sensitivity increased to 69% with a 1.45 FP per patient with CADe. Evaluation with CADe initially increased sensitivity by 14% with an increase of 0.08 FP per patient. Later CADe additions were associated with a 2% increase in sensitivity and 0.22 FP per patient. 71% of reader time was spent on the initial search and 29% with the subsequent CADe evaluation. True positive detections from CADe increased evaluation time by an average of 9.5 seconds. False negatives took an average of 8.4 seconds to evaluate. True negatives required 4.7 seconds and false positives averaged 14.4 seconds. The authors concluded that if CADe detections are ordered by likelihood of being a nodule, then an initial period of rapid performance improvement can be seen with the addition of CADe to CT interpretation. It is unclear how many true positives were presented before the first false positive and whether the differences in timing of true/false positives/negatives would remain similar in clinical practice.

Van Beek et al[7] conducted a 5-month study of real-time radiologist performance with addition of CADe on a set of 214 chest radiographs while searching for lung nodules during follow-up of known cancers. Nodule size was not indicated. Follow-up was done within 3-months to confirm nodules on subsequent scans. When digital radiographic images were read by one of three pulmonary radiologists without CADe, the authors reported a sensitivity and positive predictive value of 63.6% and 92.1%, respectively. The sensitivity increased to 92.7% ($p < 0.0001$) and the positive predictive value decreased to 89.5% with the addition of CADe. Specificity decreased from 98.1% to 96.2% (p-value not significant) when CADe was used. The authors concluded that the interpretation of lung nodules can be improved using CADe with a minimal increase in false positive interpretations.

White et al[8] conducted a study of 109 patients with 436 radiographs (in quadrants) at 4 centers. The purpose of the study was to evaluate CADe as a second reader. Nodule size ranged 4 mm to 30 mm and slice thickness varied between 0.9 mm and 3 mm. The average increase in ROC area under the curve for the 10 readers with the CADe software was 1.9% (95% CI 0.8-8.0%) from 86.7% to 88.7%. Though not statistically significant, radiologists with less experience had more improvement when using CADe compared with radiologists with more experience. Slice thickness also influenced the performance of CADe with sensitivity of 81% for 0.9 mm slices and 51% for 3 mm slices. The authors concluded that using CADe improved sensitivity in detecting pulmonary nodules.

IV. <u>Colon: Retrospective Reader Performance and Clinical Studies</u>

Summers et al 2008[9] performed external validation of an existing CADe system by comparing its performance on 104 screening patients with the performance parameters obtained in another non-polyp-enriched screening cohort in which the CADe system was trained. In the “training set” population, the sensitivity for adenomatous polyps 10 mm or larger was 93.3%, and 51.1% for adenomas 6-9 mm. There was a mean false-positive rate of 8.6 per patient. The CADe system had per-polyp sensitivities of 91.5% for adenomas 10 mm or larger and 82.1% for adenomas 6-9 mm. The per-patient sensitivities were 97.6% for patients with adenomas 10 mm or larger and 82.4% for patients with adenomas 6-9 mm. The mean false-positive rate was 9.6 per patient. In a random sample, 72.5% of false-positive findings were attributable to folds or residual feces. The authors concluded that the CADe system has a high level of performance in the detection of adenomatous polyps with CT colonography data. However, the performance of the CADe system was evaluated in a population with polyps of large enough size to undergo optical colonscopy. Most optical colonoscopy-confirmed adenomas were 10 mm in diameter or larger. Thus, these findings are less generalizable to the general population with a higher prevalence of smaller polyps and flat or sessile adenomas. In addition, the authors did not address the impact of the radiologist training and acceptance level on the performance of the CADe system.

Summers et al 2009[10] compared the radiologists' blinded visual assessment of polyp conspicuity and the detection of colonic polyps by CADe on CTC. One polyp from a CTC examination of each of 29 patients (size range 6-10 mm) from a larger cohort was grouped by the CADe system as “detected” or “undetected.” The conspicuity of the polyps was scored with a 4-point scale (0

= least conspicuous, 3 = most conspicuous). Inter-observer agreement (weighted kappa 0.38 +/- 0.15) and intra-observer agreement (weighted kappa 0.57 +/- 0.09) were fair. Polyp height was a major determinant of visual conspicuity (r^2 between conspicuity and manual measurement of polyp height = 0.38-0.56, P < .001). The performance of CADe system in detecting flat lesions/polyps is unclear.

V. Lung and Colon CADe: Research Recommendations for the Future

This memo has reviewed a number of research studies conducted to evaluate the safety and effectiveness of incorporating CADe readers in the interpretation of lung and colon computed tomographic (CT) examinations. These studies do not present conclusive evidence regarding added benefit by the introduction of CADe leading to varied acceptance of the technology into clinical practice. Although the reviewed studies indicate that cancer detection increases when using CADe, these findings are often only in novice readers and the rate of false positive results also increases. Given the current literature, it is unclear whether these negative findings demonstrate added benefit of including CADe. This is a question that must be evaluated in future studies.

The majority of published studies to date are reader studies. The data sets were often small and the CADe system was often trained on and subsequently evaluated the same data set. Thus, the reported performance was likely biased to the available data set[11] with the degree of bias increasing as sample size decreased. Due to these concerns, these studies cannot be directly compared[11].

Any study should be designed to collect information on the associated cofactors that may affect findings. These factors are listed in the table below. These studies should be conducted on large, diverse samples of patients in clinical settings. Although the randomized controlled trial is preferred, large observational studies have merit as well.

Study Variables Necessary for Full Evaluation

Factor	**Reasons for Inclusion**
Race	Cancer types and rates vary across different race and ethnic groups
Slice Thickness	Performance of CADe software is influenced by slice thickness[11] with greater sensitivity of detection in thinner slices.
Radiology Training and Experience with CT reading	Factors that affect interpretation include experience, Board Certification/fellowship training, volume and observation time
Radiologist Decision-Making	The impact of the medicolegal environment on decision-making should be evaluated including the decision-making to discount a mark made by CADe systems.
CADe Training	Due to the differences and complexities of decision-making using CADe systems, a detailed training program is necessary to improve results. Any comparative study should outline specifically the type and duration of training.

Factor	Reasons for Inclusion
Type and location of lung nodules detected	CADe systems are more effective in identifying solid lung nodules compared to ground glass nodules[11], particularly in the peripheral lung.
Type of colon polyps/lesions detected	CADe systems are less effective in identifying flat polyps and lesions[10], particularly in presence of residual fecal matter.
Baseline cancer prevalence in the population	When evaluating sample size, baseline prevalence should be used in the power analysis.
Recall rate	There is a positive relationship between recall rate and sensitivity
Observation time	Some cancer types, or the position of the cancer are only detected if the observation time is sufficient to study the film.
Adjustments to the CADe programs	CADe systems can be adjusted to balance sensitivity and specificity by adjusting the number of marks per film. This adjustment will affect any study and should be included in the evaluation of the system.
Single-Reader + CADe versus two independent readers one with CADe and one without CADe	When a study is designed using a single reader who will look and evaluate the film, make a determination and then look again with CADe and make a determination, there is the potential that both readings are biased by the knowledge that there will be a CADe available. This may cause the reader to become more conservative in the non-CADe assessment, affecting comparisons of sensitivity and specificity.
Biopsy Rate	The goal of an efficient screening tool is to minimize unnecessary exposure to additional intervention. False-negative readings often lead to biopsy. This procedure poses an additional set of both physical and psychological risks as well as additional health care costs.
Baseline sensitivity/PPV, specificity/NPV and recall rate	In order to fully evaluate the impact of double-reader or reader + CADe, it is critical to have baseline information on performance in the environment where the study is conducted
Study PPV and NPV	Although sensitivity analysis will predict the ability of the reading method to identify cancers, the true ability of the test is measured by both positive predictive value and negative predictive values since they include population parameters. Additionally, these values should be calculated on all screening cases, often they are only calculated on identified cancer cases, this may provide misleading information for interpretation.
Area under the curve or ROC analysis	Since there is a positive association between recall rate and sensitivity, a more accurate measure of improvement through CADe would be ROC analysis that would give the absolute improvement by controlling for other factors such as a change in threshold

Factor	Reasons for Inclusion
Population characteristics (including previous screens)	Patient history can affect cancer rates and the decision-making of the clinician.
Patient Safety • Anxiety • Impact of increased recall rate • False-negative rate • Number of missed cancer cases	Potential unintended adverse consequences from the use of CADe in actual practice is a critical consideration and should be explored as secondary endpoints through administration of tools for pain and anxiety and calculation of increased recall rates, additional biopsies, the impact of the false-negative rate on both patients and cost and the impact of the number of missed cancer cases.
Patient Survival	Available studies are limited by little information on long-term survival following screening. Since the goal of early cancer detection is to prevent associated death, analyses should include long-term follow-up to determine if patients develop cancer (i.e. cancer missed at screening or detection not early) and subsequently die due to the disease.

VI. Conclusions

Very little research has been conducted on lung or colon computer assisted detection (CADe). The majority of published studies are stand alone reader studies. These studies suggest that CADe does increase sensitivity of nodule or polyp detection, but there is a trade off with an increase in false positive rates as well. Further research using well designed, prospective studies or retrospective reviews is necessary.

Mary Elizabeth Ritchey, RN, MSPH, PhD, Epidemiologist
Cara Krulewitch, CNM, PhD, FACNM, Team Leader
Hui Lee Wong, PhD, MS, Epidemiologist
CDRH/OSB/DEPI

Document History:
Drafted: Ritchey/Wong/Krulewitch, October 8, 2009
Cleared: Marinac-Dabic, Date October 8, 2009

VII. References

1. Centers for Disease Control and Prevention, [http://www.cdc.gov/cancer/lung/statistics/], last accessed October 7, 2009. .
2. Shiraishi J, Abe H, Li F, Engelmann R, MacMahon H, Doi K. Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs ROC analysis of radiologists' performance. Acad Radiol 2006;13(8):995-1003.
3. Welch HG, Woloshin S, Schwartz LM, Gordis L, Gotzsche PC, Harris R, Kramer BS, Ransohoff DF. Overstating the evidence for lung cancer screening: the International Early Lung Cancer Action Program (I-ELCAP) study. Arch Intern Med 2007;167(21):2289-95.
4. Horner MJ, Ries LAG, Krapcho M, Neyman N, Aminou R, Howlader N, Altekruse SF, Feuer EJ, Huang L, Mariotto A, Miller BA, Lewis DR, Eisner MP, Stinchcomb DG, Edwards BK. SEER Cancer Statistics Review, 1975-2006. National Cancer Institute. Bethesda, MD:http://seer.cancer.gov/csr/1975_2006/, based on November 2008 SEER data submission, posted to the SEER web site, 2009.
5. Kasai S, Li F, Shiraishi J, Doi K. Usefulness of computer-aided diagnosis schemes for vertebral fractures and lung nodules on chest radiographs. AJR Am J Roentgenol 2008;191(1):260-5.
6. Roos JE, Paik D, Olsen D, Liu EG, Chow LC, Leung AN, Mindelzun R, Choudhury KR, Naidich DP, Napel S, Rubin GD. Computer-aided detection (CADe) of lung nodules in CT scans: radiologist performance and reading time with incremental CADe assistance. Eur Radiol 2009.
7. van Beek EJ, Mullan B, Thompson B. Evaluation of a real-time interactive pulmonary nodule analysis system on chest digital radiographic images: a prospective study. Acad Radiol 2008;15(5):571-5.
8. White CS, Pugatch R, Koonce T, Rust SW, Dharaiya E. Lung nodule CADe software as a second reader: a multicenter study. Acad Radiol 2008;15(3):326-33.
9. Summers RM, Handwerker LR, Pickhardt PJ, Van Uitert RL, Deshpande KK, Yeshwant S, Yao J, Franaszek M. Performance of a previously validated CT colonography computer-aided detection system in a new patient population. AJR Am J Roentgenol 2008;191(1):168-74.
10. Summers RM, Frentz SM, Liu J, Yao J, Brown L, Louie A, Barlow DS, Jensen DW, Dwyer AJ, Pickhardt PJ, Petrick N. Conspicuity of colorectal polyps at CT colonography: visual assessment, CADe performance, and the important role of polyp height. Acad Radiol 2009;16(1):4-14.
11. Chan HP, Hadjiiski L, Zhou C, Sahiner B. Computer-aided diagnosis of lung cancer and pulmonary embolism in computed tomography-a review. Acad Radiol 2008;15(5):535-55.